Beyond Infrastructure! Modelling the Scholarly Domain

Stefan Gradmann[1], Steffen Hennicke[2], Gerold Tschumpel[3], Kristin Dill[4], Klaus Thoden[5], Alois Pichler[6], and Christian Morbidoni[7]

[1]Literary Studies Research Unit, Leuven, Belgium,
stefan.gradmann@kuleuven.be

[2]*corresponding author*: Berlin School of Library and Information Science, Berlin, Germany, steffen.hennicke@ibi.hu-berlin.de
and Georg-Eckert-Institut, Braunschweig, Germany,
hennicke@gei.de

[3]Berlin School of Library and Information Science, Berlin, Germany,
gerold.tschumpel@student.hu-berlin.de

[4]Österreichische Nationalbibliothek, Vienna, Austria,
kristin.dill@onb.ac.at

[5]Max-Planck-Institut für Wissenschaftsgeschichte, Berlin, Germany,
kthoden@mpiwg-berlin.mpg.de

[6]Universitetet i Bergen, Bergen, Norway,
alois.pichler@fof.uib.no

[7]Semedia, Università Politecnica della Marche, Ancona, Italy,
christian.morbidoni@gmail.com

Abstract

In this paper[1] we will examine how the modelling of research processes in the humanities can inform the development of digital tools created for the enhancement and augmentation of scholarship. In particular, we will present the 'Scholarly Domain Model' (SDM) which provides a framework for the systematic investigation of the relation between scholarly practices and the emergence of digital practices and methodology in continuously evolving Virtual Research Environments (VRE). Our findings come from the EU funded DM2E project[2] which has been working on further developing a digital humanities collaboration environment centred around the semantic annotation tool Pundit[3].

1 Introduction

Over the last decades, the international institutions of research funding have been taking part in a process that could be described as the transition into the digital age. They have encouraged a variety of projects for the advancement of the Digital Humanities,[4] focusing on attempts to further the development of infrastructures for digital scholarship in the humanities. In Europe, for example, the European Strategy Forum on Research Infrastructures (ESFRI) has funded several infrastructure projects such as the Digital Research Infrastructures for the Arts and Humanities (DARIAH) and the Common Language Resources and Technology Infrastructure (CLARIN), which have since been brought together by the Data Service Infrastructure for the Social Sciences and Humanities (DASISH).[5] Each of these infrastructure projects have, in turn, influenced a number of others on a national, regional, or institutional level. Apart from the technical requirements of digital information and communication technology, they all have in common the desire to provide the building blocks for a sustainable Virtual Research Environment (VRE).[6]

Achieving a constellation of building blocks that is favourable to increasing sustainability is still a major challenge. This is due to many reasons,[7] among them a deficit of systematic investigation into, and hence a deficit in addressing, the actual research practices of humanists and their sustainable representation in the digital realm. For VREs, it is essential to understand the entire scholarly research process and offer applications and services which can support the corresponding workflow.

In this context, the research gap we identified and address in the current paper is the lack of a model which stresses the importance of creating a bridge connecting the analog and digital scholarly practices and, most importantly, stresses the recursive relationship between these scholarly practices and the models and applications reflecting on them. This kind of research falls within what is typically called 'digital humanities' and which we understand as a community of practices, regardless of their particular materiality. We therefore believe that in order to be able to discuss the 'digital humanities' in a way that goes beyond simply discussing infrastructure so that the aforementioned challenge can be overcome, we need to start from a modelling process that allows for the systematic and theoretically grounded building of bridges between practices of humanist research approaches in both the analogue and digital world.[8] In this paper, we discuss this undertaking and propose a multi-layered model that exemplifies the constituents of our modelling endeavour, which we have labelled the Scholarly Domain Model (SDM).

The SDM has been devised based on the assumption that understanding what John Unsworth (2000) had originally proposed in terms of Scholarly Primitives is central to any such approach at modelling the digital scholarly domain. Unsworth's Primitives are understood as 'basic functions common to scholarly activity across disciplines, over time, and independent of theoretical

orientation' (Unsworth, 2000). Like other models since, the SDM takes up the notion of Primitives and develops them further. Based on analysing and observing the practices of digital scholarship, we are endeavouring to acquire a better understanding of the requirements for instructing the development of sustainable infrastructures that enable scholars to harness the potential of digital technology and hence to develop appropriate digital methodologies. This requires to proceed beyond the establishment of static models to the iterative and continuous activity of modelling.[9] For this reason, the SDM is conceived as an explicit but not definite set of the constituents of the domain of digital scholarship in the humanities. In his talks Manfred Thaller has repeatedly stressed that the focus on the controversy of the 'digital humanities' should focus on the scholarly practices in the digital humanities and in particular their prerequisites and should not be predominated by arguments about labelling (cf. Thaller, 2013; McCarty and Short, 2002). Therefore we believe that modelling is the goal, not the model, and that the process should be conceived to be open and integrative.

In this regard, Linked Data standards[10] such as the Resource Description Framework (RDF), Resource Description Framework Schema (RDFS), and Web Ontology Language (OWL) constitute a well suited means for the development of the SDM, because they allow the process of modelling to be iterative and continuous since the graph of semantic statements created is extensible. As we will see, this is also an instance of a still uncommon and emerging way to think of Linked Data as an art with epistemological implications for the practice of modelling the domain of digital scholarship in the humanities (cf. Oldman et al., tbp).

Our research has been carried out by the EU funded project Digitised Manuscripts to Europeana (DM2E). One of the project's main activities has been

working on further developing a digital humanities collaboration environment which is built around the semantic annotation tool Pundit originating from the SemLib project. Pundit along with additional modules enables scholars to work with digitised manuscripts in the Linked Open Data (LOD)[11] Web. The development of this collaborative research environment and the modelling process of the SDM have partly informed each other. The results of DM2E are intended to contribute to the emerging digital, networked and distributed environments, well beyond traditional working paradigms in the scholarly culture of the humanities. The SDM plays a pivotal role in this respect as a framework for better understanding scholarly research practices and the ways digital working modes might evolve in the future.

Starting from the Scholarly Primitives by Unsworth (2000), the SDM was further constructed and refined by analysing the research literature and related models, which will be discussed in the following section. Furthermore, the conceptual input has been subsequently revised and supplemented by empirical evidence collected through a series of interviews with scholars and researchers from the humanities, and experiments using the Linked Data annotation environment Pundit.[12] Finally, the work on the SDM has continuously been monitored and counselled by the Digital Humanities Advisory Board (DHAB) where DM2E has brought together scholars of the digital humanities in Europe.[13]

In the following section 2, we further motivate our research and discuss the wider context of related research on which we built our model. Section 3 offers a detailed description of our proposed Scholarly Domain Model. Section 4 provides an outlook on how the model and the modelling could facilitate and support the development of sustainable VREs for scholarship in the humanities.

2 From Infrastructure to Modelling the Scholarly Domain

First, we will introduce the wider research context of our work on the Scholarly Domain Model (SDM) starting with the observation of the predominant focus on infrastructure in a lot of digital humanities projects. Then we will present related research literature and similar modelling efforts.

Infrastructure (Cf. Atkins et al., 2003) is required in order to enable advanced collaborative approaches of scholarly work in digital and network based environments. Thus, attempts currently under way to make such infrastructures available are essential, as described by Rockwell (2010) from a North American perspective. Most of these efforts have their roots in the National Science Foundation (NSF) initiative[14] that led to the foundational 'Atkins-Report' (cf. Atkins et al., 2003). This report introduced a layered vision of the way technical research infrastructures are related to each other (cf. figure 1).

This 'mother of all eScience layer cakes' introduced the hitherto canonical division between the blue area of supporting cyberinfrastructure and the white area of discipline-specific applications. Most initiatives following this report were to focus more or less exclusively on the cyberinfrastructure layer[15] such as the report on 'Our Cultural Commonwealth' (Unsworth and et al., 2006). The model of thought introduced by this report has also been adopted in Europe such as with the e-Science initiative[16] in the UK or the German D-Grid[17] initiative.

An important exception to this exclusively infrastructure driven position was the Bamboo project[18] which included work well beyond building of infrastructure. For instance, the Bamboo project delivered a report on scholarly practices (Bamboo, 2010) derived from extensive workshops with humanists, an approach that we have partially applied in our own research. Other European and national

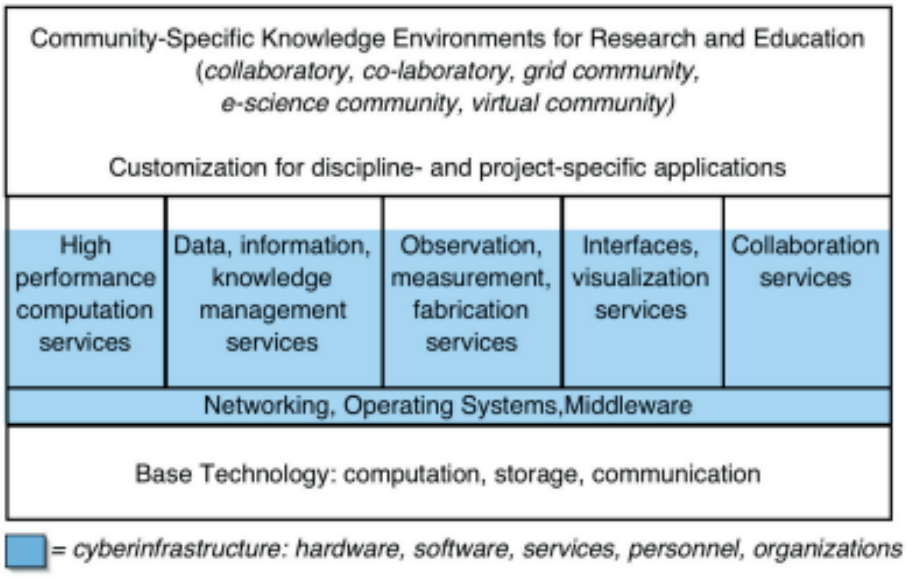

Fig. 1 The layered vision to technical research infrastructure from the Atkins-Report.

initiatives initiated a shift from exclusive infrastructure driven positions to content-based focus in the digital humanities. For example, Europeana[19], as an attempt to make representations of massive amounts of cultural artifacts available on the Web, certainly focuses much more on content than infrastructure, similar to the French humanities research platform Isidore.[20] Still, despite these exceptions, the overall tendency even in the European initiatives is mostly centred on infrastructure.

Infrastructure is not sufficient in itself if we really want to provide the tools and services the researcher needs and will use in the digital, network based environment of the Web, and, in the long-run, want to step beyond emulating traditional scholarly practices. Rockwell (2010) in his section on the 'Dangers of Infrastructure' pointed out that, when building an infrastructure, we need to be aware of two major pitfalls: Neither are research infrastructures research 'just as roads are not economic activity', nor should research infrastructures become

an end in themselves, where 'to sustain infrastructure there develops a class of people whose jobs are tied to infrastructure investment.'

Quite some research has been contributed on the issue of formalising Scholarly Activities over the past decades. Here, we do not present an exhaustive or even extensive review but only a small selection of some of the more recent and essential literature about Scholarly Primitives and related concepts. The SDM has been created starting from and based on this selection.

Unsworth (2000) conceptualised the Scholarly Primitives as basic functions which are common to any scholarly practice in the humanities independent of discipline, theoretical orientation, or era. He suggested seven recursive and interrelated Scholarly Primitives - discovering, annotating, comparing, referring, sampling, illustrating, and representing - which he saw as the basis for tool-building enterprises for the digital humanities. Since then, Unsworth's Scholarly Primitives have been often utilised and further revised. And as John Unsworth acknowledged in an interview almost a decade later, his list of scholarly Primitives is certainly not definitive (cf. Unsworth and Tupman, 2012). Subsequent research shows that there is no agreement on the exact definition or scope of Scholarly Primitives. However, the approach of using Scholarly Primitives or similar concepts appears to be a valuable and accepted means of structuring and conceptualising the scholarly domain or aspects of it. Therefore we decided to use Unsworth's conceptualisation of the Scholarly Primitives as a starting point for our own Scholarly Domain Model.

In their 'activity centric approach' Palmer et al. (2009) revised Unsworth's rather static notion of Scholarly Primitives by grouping them into 'scholarly information activities'. This approach stresses the vivid character of research and the role of information in the scholarly domain where Primitives form the basic building blocks of larger scholarly information activities. Based on an

extensive literature review, Palmer et al. (2009) identified five core scholarly information activities - searching, collecting, reading, writing, and collaborating - each of them containing several more granular Primitives, some of them being 'cross-cutting', which means they can be applied to any Scholarly Activity. Furthermore, this study indicates that the Scholarly Primitives and Activities exist universally in both the 'sciences' and the 'humanities' although in different weighting. We took a similar approach, however, in our model the Scholarly Primitives are specialised into more granular Scholarly Activities. Also, while Palmer et al. (2009) only mention different kinds of 'stages of a research project', we embedded Primitives and Activities into a wider context of a process model for research activities.

Whereas Palmer et al. (2009) based their work on an extensive literature review, Brockman (2001, p. 4) conducted early empirical research on how 'humanities scholars think about, organise, and perform their research' and the ramifications for tool building enterprises. Their study suggests four general and intertwined categories of activity: Reading, networking, researching, and writing. They conclude that such analysis of the humanist's research process constitute essential input to the development of digital tools for the humanities. In 2010, the Bamboo project performed a series of workshops with practitioners from the digital humanities in order to examine scholarly practices. They mapped their findings to the ones of Unsworth (2000) and Palmer et al. (2009). Their aim was to provide a conceptual framework for tool-building enterprises in the digital humanities. The 'Scholarly Practice Report' (Bamboo, 2010) and the recordings from these workshops are a rich source which helped us to devise the initial Scholarly Primitives and Activities for the SDM.

Apart from the research strand opened up by John Unsworth another relevant perspective is provided by the notion of the 'methodological commons' introduced

by Anderson et al. (2010), on the basis of McCarty and Short (2002). They sketched out an intellectual map which is meant to be a vivid means for mapping out the field of digital humanities. This map is intended to provide a starting point for a framework which may visualise the complex interrelations and interactions between the different disciplines, source materials, methods and technologies involved in scholarly practice of the digital humanities. In the end, and very similar to our conception of the process of modelling, the activity of mapping out the field of digital humanities has to be thought of as a continual process that is the point meant to spark off debate and to ever evolve the diagram further. Anderson et al. (2010) combined the methodological commons with the Scholarly Primitives in order to create a conceptual framework for a tool-building enterprise for the digital humanities in DARIAH. They also stress that those Scholarly Primitives should be extended beyond textual content and consider the Primitives mainly as a means of communication and explanation what traditional research activities digital tools actually enable. Similar to the methodological commons we aim at creating an integrative and vivid model of the research process but focus on its common and discrete functions and its social aspects.

Benardou et al. (2010) probably came closest to our intentions. As part of the DARIAH-EU preparatory work they devised a conceptual model of scholarly research activity which is expressed in terms of the CIDOC CRM.[21] They do not propose a comprehensive list of Scholarly Primitives or Scholarly Activities but, building upon and extending the CRM's notion of activity, show how scholarly primitives could be operationalised as properties connecting research activities with information objects and propositions, i.e. including argumentation structures. Their proposal goes beyond being a framework for categorising tools but also aims at capturing results from empirical research on scholarly research activity (also cf. Oldman et al., tbp).

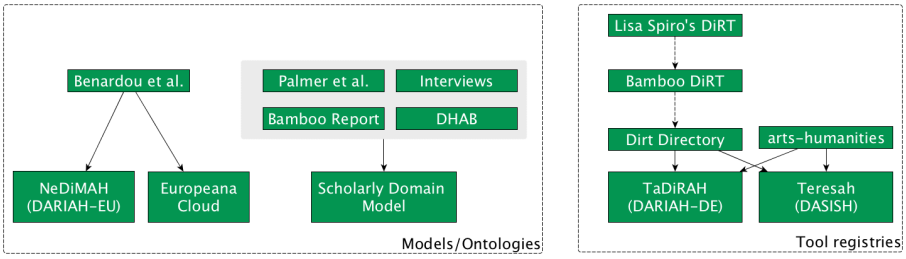

Fig. 2 Models and projects related to the Scholarly Domain Model.

The Scholarly Primitives and especially the Scholarly Activities are primarily based on the work by Unsworth (2000), Palmer et al. (2009), and the project Bamboo (2010). Whereas we started our research on this basis, utilising concepts of Activities, as well as their respective terminology and description, the concepts as they have been included in the 'Appendix: Scholarly Activities' have been subsequently appropriated for the Scholarly Domain Model. As a result of this process, some of the Activities have been substantiated, eliminated, revised and renamed or inherited as they were. We emphasise that this liste attempts to be explicit but not definite and demands to be further appropriated for future application.[22]

In the past few years, several different approaches to classifying tools and methods have emerged, some sharing the same aim as the SDM, some concentrating on being registries of existing tools. Figure 2 shows the interrelation between the different endeavours.

Although being mentioned as an inspirational source in most of the recent literature, there is no taxonomy that is directly derived from Unsworth (2000). However, it serves as an anchor for all the projects mentioned here. Figure 2 shows the genesis of currently active projects, demonstrating that there is a difference in the aims of the taxonomies. Some consider themselves to be mere tool registries while others, like the Network for Digital Methods in the Arts and

Humanities (NeDiMah)[23] and the SDM aim to describe the scholarly research practices as a whole.

The Digital Humanities Taxonomy Group[24] develops the Taxonomy of Digital Research Activities in the Humanities (TaDiRAH). The rationale, as quoted from the group's GitHub account[25], is to help endeavours to collect information on DH tools and methods. The top categories are modelled after the phases of a prototypical research process and contain more specific methods. It is pointed out that the taxonomy is not meant to cover all the methods that might exist in DH, but concentrates on a set of methods that are widely used. In addition to the category of activities there are two lists: techniques and objects. Techniques (e.g. 'Brainstorming', 'Searching', 'Encoding') specify how an activity (e.g. 'Visualization') is actually performed while the list of Objects (e.g. 'Metadata', 'Persons') is a list of objects that the technique can be applied to.[26] Both these lists are open and might change over time. From a theoretical background, the taxonomy bases itself on the concept of Unsworth's Scholarly Primitives, as well as 'the idea of a multi-stage scholarly workflow or research lifecycle.'[27] Also, the taxonomy separates research activities from research objects.

Another tool registry that is currently being developed is the DASISH Tools Registry, or TERESAH (Tools E-Registry for E-Social science, Arts and Humanities).[28] This will remain a pure tool registry and ingested data will come from both arts-humanities.net[29] and Bamboo DiRT.[30]

In contrast, NeDiMAH strives to build a formal ontology for the digital humanities including a classification and a shared vocabulary.[31] It is still in active development by DARIAH-EU's VCC2. In a presentation given at Luxembourg DH conference, December 5th 2012, Lorna Hughes stressed the usefulness of the project as it would 'formalize and codify the expression of work in DH' (Hughes, 2013), meaning also that an endeavour like this could

help and produce a common nomenclature in the digital humanities and that the use of DH methods would gain a greater academic credibility when being grounded on a theoretical basis.[32] NeDiMAH and TaDiRAH are two closely related projects, coming from the DARIAH-EU and DARIAH-DE[33] contexts respectively. Whereas TaDiRAH has a very practical approach, NeDiMAH tries to address the formalisation and classification of methods in the field. It is planned, however, to integrate TaDiRAH into NeDiMAH at a later point. The efforts in Europeana Cloud[34] are also linked to DARIAH-EU's VCC2 'Research and Education Liaison' and through this to NeDiMAH.[35] One of its ambitions is to contribute to the future Europeana Research platform. The most important report here is Deliverable 1.2 of the Europeana Cloud project (cf. Benardou et al., 2013), a desk research on the current situation of digital research practices, tools and scholarly content which gives an extensive overview over current and past studies. They conclude that even though the use of digital sources and tools has become more common and that methods in the digital humanities reflect on this phenomenon, there is still the need to support the building of infrastructures by more research on the way scholars of the humanities interact with the digital domain.

The literature and models we presented above provided us with valuable input for important categories and the overall design of our modelling approach. However, we found that these models lack a perspective that we consider important for our purposes: Although humanities exhibit an increasing drift into the digital, the major part of the scientific community is not using dedicated digital humanities tools. Rather, scholars rely on well-known but not necessary the best suited software. Thus, maintaining tool taxonomies and classifications of digital methods are necessary, but not sufficient steps on the way to modelling and supporting scholarly work as a whole.

The SDM which we are going to present next proposes a non-static model whose constituents cover the analogue, traditional activities of the humanities and put them into a general, integrative model of research that also considers the digital context.

3 The Scholarly Domain Model

The model we are proposing consists of four different layers of abstraction which will be described in the following sections. These layers are Areas, Scholarly Primitives, Scholarly Activities and Scholarly Operations.

The Areas represent the general stages of scholarly work, whose central point is of course research, but which also covers aspects of a circular workflow and surrounding, contextual Activities like social and administrative aspects which influence the research process. The next layer consists of the Scholarly Primitives that form the most abstract description of scholarly practices in the model. The Primitives are located mainly in the Research Area, but extend also in other Areas. The motivation for these Primitives is what we think the most simple description of the research process: Interpretative Modelling, Exploration, Aggregation, Augmentation and Externalisation. The third layer consists of Scholarly Activities, a set of categories for describing possible research processes. The categories we propose are still generic and not domain-specific constituents. In contrast to the Primitives, the Activities refer to particular yet generic parts of the research process which, in principle, may occur in any sequence or constellation. Thus Activities do not have an exclusive or definite subclass relation to Primitives but may be seen as relating to one or more Primitives based on the particular context of their application. The Scholarly Operations form the most concrete layer of the model. On this level, the Activities are viewed through the lens of a specific application scenario, i.e. including other

Fig. 3 Layers of Abstraction.

constituents of the SDM such as the Actors, that perform Scholarly Activities within their respective research process, determined by a Social Context as well as the the applications and material at hand.

We chose this layered division of the model as shown in figure 3 in order to be able to adapt the model to a number of possible applications during the modelling process. With this framework, the scholarly domain can be modelled on four different levels where the first three provide a systematic and structured vocabulary for the analysis of the scholarly domain while the fourth one is concerned with their respective observation in practice. The potential benefit of this is that the model, in particular seen as a modelling process, is better suited to react to the requirements of the continuous development of research infrastructures as well as of the scholars, those infrastructures are developed for. Recursion, or the facility to integrate what is 'missing' is crucial for the model to be easy to apply as well as to adapt. This recursive and adaptive modelling process can be driven by the use (as of now) of RDF to make components interact with each other on data level.

The focus of the representation within the SDM, described in the next section, is on the scholarly practices represented in the four layers of abstraction. Other pivotal constituents of the domain, such as the scholars, the Actor, or the representations of the objects of their research are not included.

3.1 Areas

The five proposed Areas - input, research, output, documentation and social context - form the uppermost and integral part of the Scholarly Domain Model with the central Area being Research, as shown in figure 4. We have chosen to start here, because it reflects the cycle of scholarly work with its different phases of gathering input for research, the act of dealing with the input and the externalisation of results. Hence the two additional Areas Input and Output. Furthermore, we cannot examine the research process in isolation, but need to address its context too, in particular the the Social Context and Documentation.

The arrows in the figure imply a sequential grouping where one Area probes into the next one. The input flows into research, research manifests and is being condensed in that process as output. The output, either intermediate results or final results, then serves again as input in another iteration of the research process.

Input covers all Activities and objects that deal with the exploration and aggregation of material that will be used for research. For example, Activities, that range from Searching the Web or Browsing library shelves, and the excavation at an archaeological site to the Selection and Assessment of objects relevant for the Research.

The arrow of input protruding into the Area of research shows that these processes of exploration and building of the corpus could already be considered to be parts of the research Area. The term corpus in our model denotes any information object the scholar collects or otherwise aggregates for the purposes of research including personal collections of research data. It is the representation of the objects of research in the SDM. The research process is not linear which leads to iterative modifications of the corpus. For example, in a later stage, when doing research properly speaking, a scholar might discover that elements

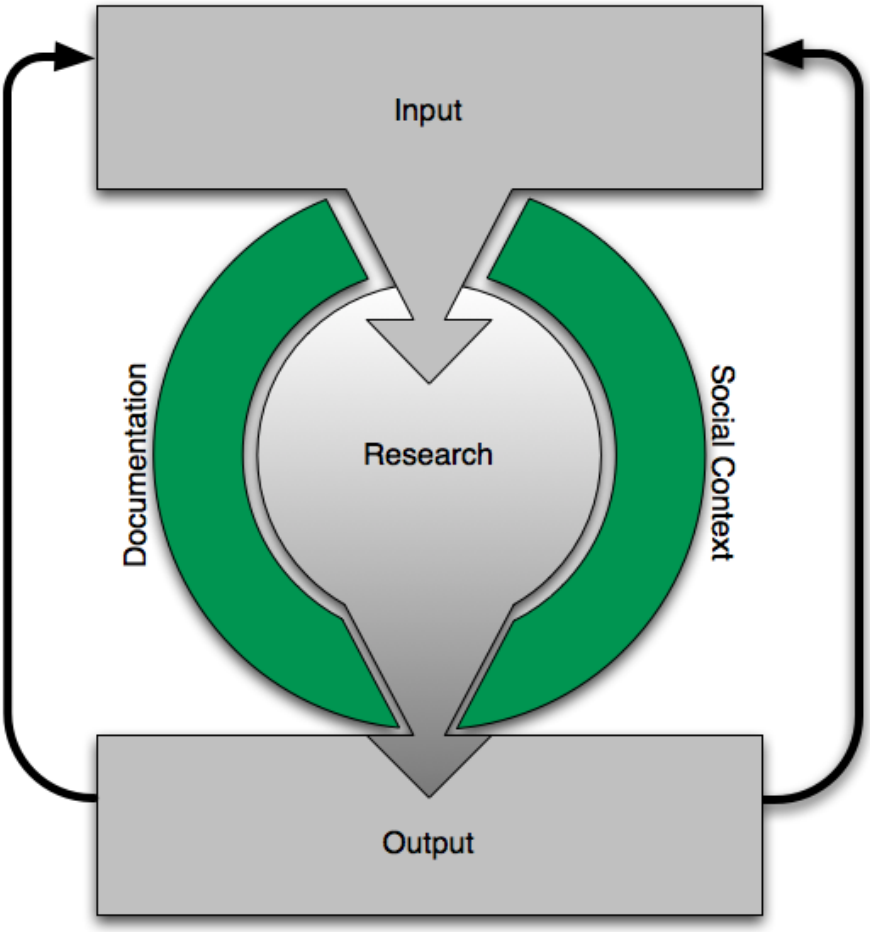

Fig. 4 The five Areas of the SDM.

are missing from her corpus which she then needs to adjust by going back into an phase of exploration.

Two additional sources serve as an input to the research process properly speaking as they exist prior to this research: Referential Data and Referential Structures, both of which are explicitly shown only in figure 4. Referential data are, for example, dictionaries that would be useful for someone reading an 18th century political tract, and wanting to see other contexts from this period in which a certain word (e.g. 'liberty') is used. Another example are Semantic Web ontologies and Linked Data resources as for instance used in the Research Space project[36] or in the Isidore environment.[37] On the other hand, referential structures such as grammar type resources, rule systems and others preexist the actual research but here again are placed in the interfacing area since – as we will see below – interaction with these in the sense of corpus contextualisation is one of the first steps in research.

Before considering the research Area in detail, the core of this layer of the SDM, we will first have a look at the output Area as well as the contextual Areas, social context and documentation.

Leaving the central Area of research as a black box for the time being, final and intermediate results of research are shared and disseminated as Output. Information that has been refined during research is now being externalised as a stable and citable information object, irrespective of its material carrier, that becomes subject to reference for either private use, limited sharing within groups, or general publication. The potential of this externalisation to enter subsequent iterations of the research cycle is assessed. Output typically entails also a change in availability of these research results: what has been kept in seclusion until now or has been shared with only a few colleagues and members of working groups is released to the public.

Note that – as was also the case with Input – there is overlap with research properly speaking. The discursive and technical organisation of research output is to some extent determined by the way it will be published at a later stage and vice versa. And this overlap may be significantly larger in humanities scholarship as compared to the so called empirical sciences: As was shown in Gradmann and Meister (2008), research and results in the empirical sciences can be considered to be totally disjoint, as in the case of an experiment and the paper reporting on it, whereas in the humanities there is a tendency for publication format, research corpora and scholarly discourse to be highly intertwined.

Furthermore, the bulk of the output may well come from research, but the social context and the documentation can certainly be considered relevant sources of output as in the cases of published citation analysis or project reports. Thus, the central Area, research, is additionally highlighted by the two remaining Areas, documentation and social context, depicted in green in figure 4. These two Areas form the context in which research is embedded.

The Area of Documentation reflects on the fact that research involves the externalisation of a form of meta-discourse to create accountability, transparency and the ability to retrace the single steps of research.

This may include informal exchanges related to research progress and also formal reports that need to be given to funding agencies. Taking the digital humanities as an example, correspondence via email about research, the keeping of notebooks and comments made when checking in source code into version control is an important form of documentation.

Also, it facilitates the interpretation of research processes and results over time in creating a narrative context – which itself can become subject of research. Furthermore, there is a need for a discourse about research itself in Science Studies as well as in the history of science, and documentation provides the

material basis for this discourse.

The Area of Social Context reflects on the fact that research is determined by the socio-historical situation in which it occurs. This includes such factors as domain specific research practices, the customs of research communities as well as national and international academic cultures. The SDM accounts for this by acknowledging the existence of the Social Context and that it affects the Scholarly Activities and Scholarly Operations carried out by researchers. A fact that is overlooked easily by ongoing infrastructure projects: social influences such as research practices are important, especially in interdisciplinary endeavours. And, as already stated for documentation, it can inform the meta-discourse about the research process. Rules, control and incentives are key notions in this social context of VREs.

The importance of including the Social Context in the SDM can be made clear by the example of citation which determines the way a citation looks like in a publication such as the conventions in a discipline or style guides by publishers, who is actually cited, often caused by political aspects or regarding the career of the author, and what is cited.

Research is not as exclusively content oriented and content driven as many of us tend to think. Many aspects are often motivated or constrained by the social context, leading to research results being a complex amalgam of content and its apprehension by the scholarly community. Collaboration with others in the research process is a sensible issue in this respect, requiring highly granular and controllable data privacy settings. What is secluded in one moment may be shared with a close community in a second moment and after publication inversely would require extreme visibility in order to obtain references, citations and crediting by the scholarly community.

The Research Area focuses on those elements which constitute scholarly

research at its core. The other Areas – Input, Output, Documentation and Social Context – share various of their elements and interact with the constituents and Activities in the Research Area. We will first describe how the various Areas interact with the Research Area.

Apart from the Primitives present on this layer there are three additional constituents in the Research Area, as shown in figure 5. The most important of those is the corpus, the body of sources the scholar decides to work with. In this abstract model we refrain from specifying anything particular about the consistence of the corpus, but it might contain any sort of information objects, and any sort of data that are manageable by machines including their metadata and data model. Adding objects to and removing from the corpus might be as simple as bookmarking a page in the Web browser or returning a book to the library or as hard as excavating the ruins of a Roman temple. The other two are, as already mentioned above, referential data and referential structures. They are auxiliary entities which are used to contextualise elements of the corpus, for example linking to authority files or Linked Data resources, or to embed one's own research into a broader context, like a theoretical framework. This shows that the various Areas are not strictly separated but are fading into each other.

Within the SDM, the research process begins with creating the corpus and contextualising its elements using referential data and structures. The basic process underlying contextualisation, and for that manner conceptualising, within the framework of the SDM is the Primitive interpretative modelling, as it is underlying all scholarly research, representing the process of 'understanding' the corpus and its constituents for the purposes of the research process. We assume that any kind of Scholarly Primitive and Scholarly Activity is always grounded in this Primitive. Interpretative modelling forms therefore the core of the research Area. In addition, we propose at least four additional Scholarly

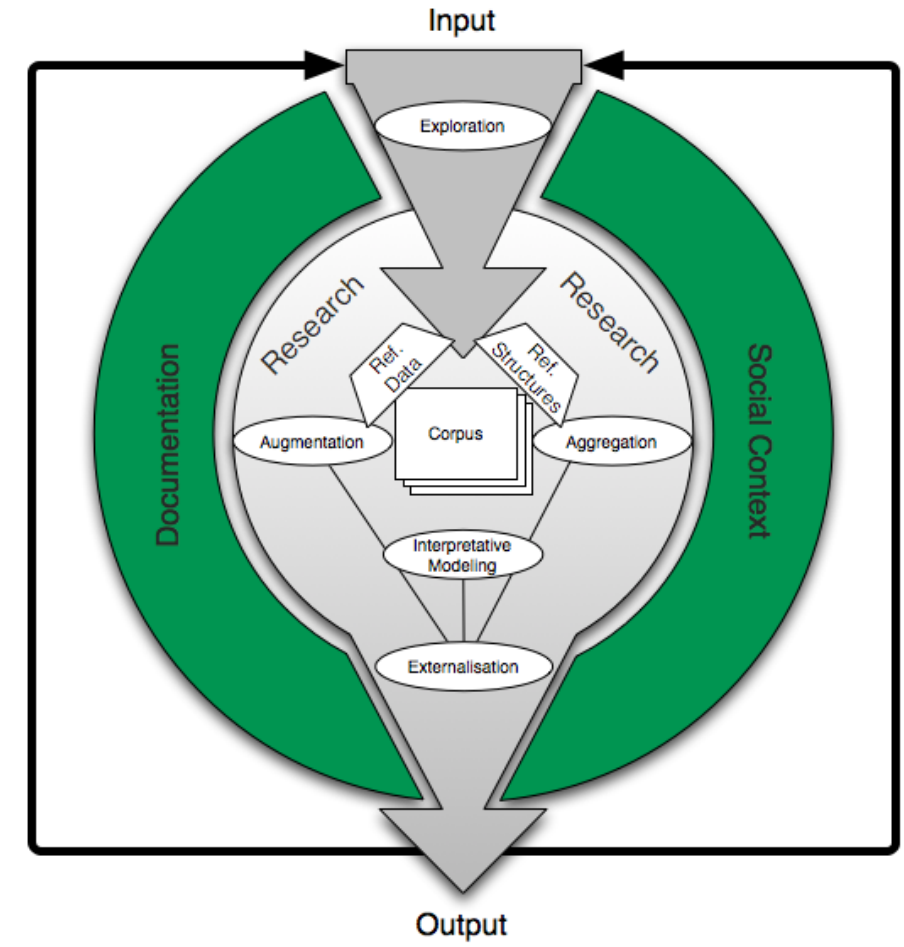

Fig. 5 The Research Areas of the SDM.

Primitives: exploration, aggregation, augmentation and externalisation.

In order to show how the other Areas are related to research, consider the following example of a researcher analysing the works of Ludwig Wittgenstein. This imagined researcher would first of all explore and determine the input of her research and build a corpus of relevant sources and articles, possibly from the Wittgenstein Repository and Wittgenstein Source provided by the Wittgenstein Archives Bergen (WAB), and possibly by utilising a faceted browser, the Wittgenstein Ontology Explorer. The faceted browser helps to iteratively focus and zoom in onto the sources and metadata, and restrict the corpus to a selection of items relevant for the specific research question, e.g. whether and which visual analogies occur in the context of Wittgenstein's remarks on the nature of philosophy. With the ongoing research process, she would keep her working group, for example other Wittgenstein scholars that have made their Pundit notebooks public, updated on the progress through sharing it with them in her own Pundit notebook (social context). She would document any additional findings in separate Pundit notebooks and would inform her university on the progress of work (documentation). Once there is an accumulation of valid and valuable research results, she would generate output by for example presenting the results at the annual international Wittgenstein Symposium in Kirchberg. Ideally, she would also publish her aggregated set of research data extracted from the corpus – together with the processing methods she had used in her research for others to use as input for subsequent research projects.

This example also makes clear that the picture should not be read as a static arrangement of components nor as their linear succession of a sequence of Input (start) to Research to Output (end). The arrows pointing from the output to input are meant to visualise a circular process in which the output of one iteration can be and typically is input for the next. To really comply with the

complexity of the research process its recursive nature might even be organised as a spiral in order to indicate progress instead of eternal repetition.

Next, we will have a closer look at the Scholarly Primitives which we think represent the most basic constituents of any humanistic research process.

3.2 Scholarly Primitives

The Scholarly Primitives constitute the most generic and principal parts of any research process in the humanities and form the second most abstract building blocks of the model. They facilitate the initial description of research processes in a very abstract but still generic way and constitute a basis for proceeding to more specific representations.

The basic set of Primitives that we propose are interpretative modelling, exploration, aggregation, augmentation and externalisation. These are inspired by the work of Unsworth (2000), but have been refined further by the study of literature (cf. section 2), interviews that we conducted and the counsel of the Digital Humanities Advisory Board (DHAB) as well as observations during experiments with Pundit.

As stated above, the Primitive Exploration is located primarily in the Input Area and thus occurs before or in-between research cycles. Exploration is about serendipitously navigating networks of related information objects that will lead to the creation of the corpus. The corpus that is then gradually built up will be the object of Activities like direct searching, browsing and rearranging, but these kinds of tasks are situated on a less abstract level of the model on the level of Scholarly Activities. In this regard, the process of direct searching must be differentiated and can be seen as a particular case of exploration.

As stated above in the section on the Area of research, Interpretative Modelling is the basic constituent Primitive that makes up the central element in

research and serves as a hub for the other Primitives. The process of 'understanding' is what it represents at its core as it revolves around the corpus by contextualising and conceptualising its elements, successively re-aggregating and re-arranging it to finally be able to externalise ideas are the core Activities here.

The Primitive Aggregation consists basically of arranging or rearranging the corpus elements. Filtering and sampling are examples of such aggregation activities that typically result in rearranged elements of the corpus such as, for instance, the pages of a digital edition arranged according to their relevance for the research process or the pieces of a vase found in an excavation arranged for their reconstruction.

Augmentation adds to the elements of the corpus. Annotations and comments are typical examples, but also context links added to the corpus elements. Such augmentations are results of research in their own right, even though their potential for publication is controversial among scholars in the humanities.

Finally, instances of Externalisation such as critical texts, textual interpretation or visualisations have to be produced to make the results of interpretative modelling and therefore the research process 'readable'.

This list of Scholarly Primitives, and also the list of Scholarly Activities which will be discussed in the following section, are taken from our research undertaken in the context of our project (cf. section 2). Some scholars may find that our Scholarly Primitives do not capture or capture incompletely what they consider to be the Primitives of their own field, or they may feel uneasy with the terminology. In the context of our perspective on modelling we see the SDM as an abstract proposal that provides a domain-independent framework open to further iterations of adaption and specification for the application in more domain specific scenarios. The particular Primitives and Activities are explicit, but not definite.

To continue with our example from the previous section we might take a more specific look at the research progress. In the phase of exploration, the Wittgenstein scholar would browse through the Nachlass on Wittgenstein Source and secondary literature in the Wittgenstein Repository, and might also use catalogues and finding aids for building up the corpus such as the Wittgenstein ontology[38] provided by the WAB (aggregation). The corpus is enriched (augmentation) by linking the sources to the Wittgenstein ontology (referential structures) and by looking up references in a lexicon, for example the Glock Wittgenstein Dictionary[39] (referential data), and hereby contributing to augment the original ontology further. The central part of research, the interpretative modelling will take its course. Finally, as an act of externalisation, the results are written up as an article to be submitted to a journal, for example the Open Access Nordic Wittgenstein Review.[40]

3.3 Scholarly Activities

The Scholarly Activities constitute the most concrete of the abstract layers of the SDM. As with the Scholarly Primitives, they are primarily based on or have been taken from the work by Unsworth (2000), Bamboo (2010), and Palmer et al. (2009), and have been partly further revised by the development and research conducted in the context of the DM2E project including interviews and a series of experiments with humanists (cf. section 2 above). We propose 25 different Scholarly Activities.[41] We do not consider this list to be definitive, in particular in terms of their number or the labels used for the Activities as well as the scope notes used to describe them. Nevertheless, since many of the Activities on the list are common in the literature as well as in the scholarly work in the humanities, they can be regarded as a recommendation. Despite the fact that a list like this might be subject to further specification for concrete application

scenarios, we want to emphasise that the observations made during our research are not exhaustive. We encourage more systematic work on these Primitives, Activities and their ontological formalisation.[42]

Despite the fact the Activities are conceived to be more specific than Primitives, the SDM does not consider their relation to be strict or hierarchical and that it is possible that each Activity can be related to one or more Primitives.[43] The difference between Activities and Primitives can be found in the different layer of abstraction used for the analysis and description of a part of the research process. As previously mentioned, Primitives and Activities typically materialise as sequences, that iterate, or in specific constellations. Therefore, it may appear difficult to determine the relation between Activities and Primitives while observing them as one part of a research process or another. For example, as mentioned earlier, there is some form of interpretative modelling involved in all scholarly research practices.[44] Furthermore Activities and Primitives may also be part of one of the Areas social context, documentation, input, or output.

Since the proposed list of Scholarly Activities, contains 25 items, we refrain from discussing each one individually, but we discuss two Activities as examples: annotating and contextualising. As described in the scope notes (cf. appendix), Annotating is considered to be the Activity of 'adding any kind of notes or markings to elements of the corpus.' This results in enriching an element of the corpus with additional data, for example, this could be a note written in the margins on the page of a book or – as will be seen below – the markup of an electronic resource using RDF triples. The creation of an annotation is accompanied by a series of other Activities. As annotating itself can also be an act of writing, what is being written down can be an act of translating, contextualising or comparing. At the same time, this piece of information is another element that is being added to the corpus. Another important and far-reaching Activity is

contextualising which we already encountered earlier. This one would be related to the interpretative modelling Primitive, but connects items of the corpus to referential data and referential structures. Thus, relationships are created either between objects that are part of the corpus, but also between objects in the corpus and external sources. As before with annotating, contextualising also resonates in other Activities. A part of the contextualisation is often a reference or a link to another source, so that in this case referring/linking is an adjunct Activity.

3.4 Scholarly Operations

The Scholarly Operations are the concretisation of the Scholarly Activities for a specific application scenario. This concretisation therefore depends on the purpose or the focus of the observation that is intended for the respective scenario. An Activity could be translated into a variety of Operations, with a variety of different constituents, for example regarding citation, as an instance of the Activity referring/linking, the focus could either lie on quantitative aspects of citation behaviour or on qualitative aspects such as different types of citation relations. Thus, for observations of Scholarly Operations focused on the quantitative aspects the actors and a model of their social context, are imperative constituents. Whereas the observation focused on qualitative aspects, might require different constituents regarding the linguistic classification of citations.

In addition to that, as each scholarly discipline or community has its own specific requirements, concerning the applications and the conventions of their scholarly practices, further constituents would have to be determined for the specification of application scenarios. The scholarly practices of the Activity comparing, for example, vary greatly in different disciplines. When comparing two or more different versions of a Middle High German manuscript, the differences

between the versions can be computed and visualised by software and might serve as a basis for a critical edition of the manuscript, provided the input texts contain appropriate markup. Scholars of Art History or similar disciplines dealing mainly with images will have other needs and means to assess the differences or similarities of the objects in comparison.

Since the focus of this paper on modelling the Scholarly Domain focuses on the description of the process rather than the description of the application of a model, the following section does neither attempt to provide a comprehensive description nor to conduct a systematic investigation on how the various abstract Scholarly Activities could materialise in concrete Scholarly Operations. Nevertheless, the next section does strive to delineate their particular relation in examples from the context of DM2E. We will discuss how the SDM and its practice of modelling could be used to instruct the development of VREs for digital scholarship in the humanities and how the use of the Resource Description Framework (RDF) as a principal data model could help to sustain its operations.

4 From the SDM to Modelling the Scholarly Domain

Despite the fact that the development of the Scholarly Domain Model (SDM) and the development of the Virtual Research Environment (VRE), with the semantic annotation application Pundit at its core, has not been systematically integrated within the Digitised Manuscripts to Europeana (DM2E) project, they influenced each other as theory and practice of modelling from the very beginning. The current version of the VRE enables various Scholarly Activities on an application level such as providing facilities for the collection and creation of vocabularies as well as annotations.

For research infrastructures to be able to sustain digital scholarship in the humanities, we believe that the scholarly practices as well as the continuous development of applications by integrating the lessons learnt through the observations of user behaviour has to be taken into account. Therefore, the SDM has not been devised to be another attempt to establish a static model but rather to instigate an iterative and continuous process of modelling. For this reason, the SDM is conceived to provide an explicit but not definite set of the constituents of the domain of digital scholarship in the humanities.

In DM2E, we conducted a series of experiments[45] in order to approach the observation of the Scholarly Operations within the framework of different application scenarios associated with the Pundit environment. The experiments demonstrate how the manifestations of the Scholarly Activity annotating as RDF vocabularies and statements vary in respect to the different research processes of interpretative research in the humanities. The implementation of the Scholarly Activities in applications is a prerequisite for systematic observation of how they specialise into Scholarly Operations in different application scenarios. In this context, the RDF data model which underlies the VRE developed in DM2E[46], is a suitable means not only to connect the Activities on a data level, for example, to make annotations explorable alongside the vocabulary used for the annotations, but also to create explicit and formal representations of Scholarly Operations in the first place. The translation of different research interests into simple annotation vocabularies represents one of the possibilities to create Scholarly Operations for observation. They operationalise, i.e. explicate and formalise, the Activity of annotating as RDF statements along with various conventions and guidelines which means that Scholarly Operations may very well consist of different constituents. Furthermore, the experiments also demonstrated that interpretative modelling is indeed influencing and present during the application

of the vocabulary, the Activity of annotating, but also during the creation of the vocabulary and the evaluation of the results, in this case through visualisation in faceted browsers. In this context, the terminology of the SDM provides a framework for systematic investigation and operationalisation of scholarly practices, i.e. their translation, again in our case into a Linked Data environment.

The translation of the Scholarly Activity annotating, as well as the translation of the research interest of the respective processes into Scholarly Operations unveiled the inherent relationship between the practice of modelling and the scholarly practices. The Scholarly Operations are mere constructs in the context of specific use cases determined by what we want to observe and what we can observe in particular research processes. As such, Scholarly Operations express and formalise what we would like to and what is possible to analyse and hence may instruct the further development of applications and methodology. In other words, Scholarly Operations are constructs of observation, and as such they serve the purpose of analysis which again serves development of tools and their application.

Secondly, the translation and application of the Scholarly Operations unveils the relationship between modelling and methodological reflexion as the research process is conducted. Scholarly Operations, in the current example the annotation vocabularies, evolve since during their application new constructs may emerge to be represented.

The experiments also suggested that the annotation acts conducted in the context of the interpretative work could be further structured into templates of several combined statements which will be reused. Such RDF-templates are one example of a first step in the direction to substantiate the process of recursion and to be able to approach the representation of the Scholarly Domain by 'modelling' rather than by a 'model'. Such templates can be modelled in

RDF as sets of triples which describe the kinds of statements involved in certain Activities. For example, a template for the Activity selecting may contain a criterion for that Operation, an actor who performs it, as well as the item which is either removed from or added to a corpus and related metadata as constituents. Since RDF allows to specialise properties and classes, communities or single users may create more specific statements within a particular RDF-template. By using such templates, we connect the abstract and conceptual level of the SDM, the model, with the concrete and explicit level of modelling and performed Scholarly Activities.

The second step in the direction to substantiate the process of recursion includes monitoring, either in the analogue or the digital realm. The latter in particular has the potential to proceed to further and to instruct the development of applications by automatising the observation of their usage for their successive analysis to adapt the applications according to the actual conduct of scholarly practice. The Patterns identified in such a monitoring of Operations can, for example, be fed back for the adaption of the aforementioned templates and thereby retain the adaptive modelling process. A potential use case has been discussed with the project Virtual and Real Architecture of Knowledge[47], a part of the project Image, Knowledge, Gestaltung[48] at Humboldt-Universität zu Berlin[49], who are planning to monitor and record all digital and analogue interaction of researchers within a laboratory and to extract and model typical patterns of behaviour. The SDM has been taken under consideration to provide an ontological framework for the representation of such patterns since it provides enough flexibility to provide a starting point for such an endeavour. The observed patterns of usage and user behaviour could be integrated into the SDM representation with RDFS/OWL, and consequently be implemented into an application such as Pundit to substantiate the monitoring, for example, of

Activities for the documentation of the respective parts of the research process. For the SDM as a framework for integration one of its benefits may become apparent, when it is taken into consideration that the extent of the automated creation of machine-processable data from monitoring Activities also impacts the potential subjects for analysis.

Nevertheless, the experiences we made during the experiments as well as the discussions regarding potential applications of Linked Data and Reasoning technologies in humanities scholarship (publication in preparation) point to the fact, that as recognisable and significant they may be as careful and delicate they have to be treated not to overestimate future developments. In all described cases, the limits of such RDFS/OWL formalisations need to be identified and kept well in mind in order not to move into the 'Artificial Intelligence Rathole', as adequately termed by Wendy Hall.[50] The aim cannot be to substitute creative thinking, as has been identified in Gradmann (2010), but to assist the scholar during the research process with functionality that, on the one hand, remains rooted in traditional and established processes but, on the other hand, also allows to go beyond using digital infrastructure for the emulation of traditional Scholarly Activity. That is why modelling is so important to be thought of as a continual and iterative process that integrates the development of the applications of digital scholarship as well as the basis, in which their use is grounded, the scholarly practices of the humanities.

5 Conclusion

The Scholarly Domain Model (SDM) has been developed in the light of a recognisable deficit in conceptual work on the constituents of scholarship in the digital humanities and a predominance of infrastructure-oriented projects in the field. The SDM provides a framework for the systematic investigation of the

relation between scholarly practices and the emergence of digital practices and methodology in continuously evolving Virtual Research Environments (VRE).

Despite the fact that the SDM has been devised in the context of applications based on Linked Data, the model is independent from particular representations and meant to be applicable as a reference model for the discussion, evaluation and development of digital research infrastructures for the humanities. The SDM allows to create representations of the workflow of digital humanists and to function as a terminological bridge between the humanities and digital applications. Only if we better understand how scholars undertake their research now and in the past and how their functional framework might be adequately translated to the digital environment, we might actually approach the emergence of new digital modes of working.

Furthermore, the SDM differs from similar approaches in so far as it approaches the scholarly domain from a more comprehensive perspective and tries to integrate Primitives of the process of scholarly work and various layers of abstraction rather than isolated acts. The model stresses the importance of recursive and continual modelling processes in order to adapt VREs to evolving scholarly practices. Then again, we believe the modelling is the goal, not the model.

Acknowledgements

The following members and associates of the Digital Humanists Advisory Board (DHAB) contributed to the development of the Scholarly Domain Model and the writing of this paper: Tobias Blanke (UK), Sally Chambers (Germany), Alastair Dunning (Netherlands), Stefan Gradmann (Belgium), Jonathan Gray (UK), Gerhard Lauer (Germany), Christian Morbidoni (Italy), Alois Pichler (Norway), Jürgen Renn (Germany), Laurent Romary (Germany/France), Felix

Sasaki (Germany), Susan Schreibman (Ireland), Claire Warwick (UK), Dirk Wintergrün (Germany)

We also would like to thank Dominic Oldman for his valuable feedback and input on the paper.

This work was supported by the European Commission (CIP-Best Practice Network) [297274].

Notes

[1]An earlier version of the current paper has been published as part of Deliverable 3.4 'Research Report on DH Scholarly Primitives' of the EU-funded DM2E project.

[2]http://dm2e.eu/

[3]http://thepund.it/

[4]We understand this term to be grounded in the basis of the translation of the German word for *Geisteswissenschaften* and not in the political sense (also cf. Gold, 2012; Terras et al., 2013).

[5]ESFRI: http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri, DARIAH: http://www.dariah.eu/, CLARIN: http://www.clarin.eu/,DASISH:http://dasish.eu/about_dasish/

[6](Cf. Candela et al., 2013), respective Research Infrastructure.

[7]Cf. the comprehensive work of the European Science Foundation as presented in Moulin et al. (2011*a*,*b*).

[8](Cf. McCarty, 2005) as well as Beynon et al. (2006) who both advocate a conception of 'empirical modelling' (cf. http://www2.warwick.ac.uk/fac/sci/dcs/research/em/)

[9]McCarty (2004, 2005) as well as Beynon et al. (2006) and further section 2 on that matter.

[10]Cf. for the following standards http://www.w3.org/standards/techs/rdf\#w3c_all and http://www.w3.org/standards/techs/owl\#w3c_all

[11]http://www.w3.org/standards/semanticweb/data

[12]Detailed reports on the interviews and experiments can be found in the Deliverable 3.4 of the DM2E project.

[13]http://dm2e.eu/dhab/

[14]http://www.nsf.gov/

[15]Some passages of the report read as if the 'Base Technology' layer was also part of the cyberinfrastructure. And some participants of the group moderated by Dan Atkins may even have wished to place the focus of cyberinfrastructure rather in this base technology area - but this does not invalidate the point made here regarding the division of cyberinfrastructure and the discipline specific application area.

[16]Cf., for example, an Humanities and Arts perspective on the e-Science initiative in the UK in Blanke and Dunn (2006).

[17]http://www.d-grid.de/ as well as, for example, http://www.textgrid.de/

[18]http://www.projectbamboo.org/about/

[19]http://www.europeana.eu

[20] http://www.rechercheisidore.fr/

[21] http://www.cidoc-crm.org/

[22] The complete list of Scholarly Activities together with scope notes and references to the original or closest descriptions of the concepts can be found in the 'Appendix: Scholarly Activities'.

[23] http://www.nedimah.eu/

[24] https://github.com/dhtaxonomy

[25] https://github.com/dhtaxonomy/TaDiRAH/blob/master/introduction.md

[26] Techniques: https://github.com/dhtaxonomy/TaDiRAH/blob/master/reading/techniques.md and Objects: https://github.com/dhtaxonomy/TaDiRAH/blob/master/reading/objects.md

[27] http://www.nedimah.eu/

[28] http://teresah.angular.dev.dasish.eu/\#/

[29] http://www.arts-humanities.net/

[30] http://dirtdirectory.org/

[31] http://www.nedimah.eu/workgroups/development-ict-methods-taxonomy

[32] Further information can be found in the minutes from a workshop in 2013: http://www.nedimah.eu/reports/scholarly-practices-research-and-methods-ontology-workshop

[33] https://de.dariah.eu/

[34] http://pro.europeana.eu/web/europeana-cloud

[35] https://dariah.eu/activities/research-and-education.html

[36] http://www.researchspace.org/

[37] http://www.rechercheisidore.fr/

[38] http://wab.uib.no/cost-a32_philospace/wittgenstein.owl

[39] Glock (1996)

[40] http://www.nordicwittgensteinreview.com/

[41] The complete list of Scholarly Activities together with scope notes can be found in the appendix.

[42] Cf. a draft version of the SDM as an RDFS/OWL ontology can be found at http://webprotege.stanford.edu/\#Edit:projectId=32a9b5a3-0781-4846-b195-980482fe54c4

[43] This is not imperative, (cf. for example Palmer et al., 2009), who strictly relate Activities to Primitives. Both modelling practices have their advantages and disadvantages. The stricter the relations are the harder it gets to differentiate the vocabulary further.

[44] The report (part of Deliverable 3.4 of DM2E) on reasoning experiments conducted with Pundit discusses an example of how interpretative modelling may materialise and be translated

into a digital, Linked Data context.

[45]Cf. Deliverable 3.4 of DM2E (to be published early 2015) for a full report on the experiments.

[46](Cf. Grassi et al., 2012)

[47]`http://www.interdisciplinary-laboratory.hu-berlin.de/en/Virtual-and-Real-Architecture-of-Knowledge`

[48]`http://www.interdisciplinary-laboratory.hu-berlin.de/en`

[49]`https://www.hu-berlin.de/?set_language=en&cl=en`

[50]At the Cultural Heritage and the Semantic Web British Museum and UCL Study Day, British Museum, London, January 2011.

References

Anderson, S., Blanke, T. and Dunn, S. (2010), 'Methodological commons: arts and humanities e-science fundamentals', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **368**(1925), 3779–3796.

URL: *http://rsta.royalsocietypublishing.org/content/368/1925/3779.short*

Atkins, D. E., Droegemeier, K. K. and Feldman, S. I. (2003), *Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*, University of Michigan Library.

URL: *http://www.nsf.gov/cise/sci/reports/atkins.pdf*

Bamboo (2010), 'Project bamboo scholarly practice report'.

URL: *https://googledrive.com/host/0B3zU098zQ8VMc2xfMUJZaWxXNWs/wp-content/uploads/Project-Bamboo-Scholarly-Practices-Report.pdf*

Benardou, A., Constantopulos, P., Dallas, C. and Gavrilis, D. (2010), A conceptual model for scholarly research activity, *in* M. Reilly, ed., 'IConference Papers 2010'.

URL: *https://www.ideals.illinois.edu/bitstream/handle/2142/14945/benardou.pdf?sequence=2*

Benardou, A., Dallas, C., Papaki, E., Constantopulos, P., Angelis, S., Baldwin, T., Lefferts, M., Hughes, L., Roberts, O., Ekman, S. and Sjögren, B. (2013), 'Deliverable d1.2 – state of the art report on digital research practices, tools and scholarly content use: Europeana cloud deliverable'.

URL: *http://www.pro.europeana.eu/documents/1414567/a0a84066-a601-4737-945e-ab63484ae804*

Beynon, M., Russ, S. and McCarty, W. (2006), 'Human computing: Modelling with meaning', *Literary and Linguistic Computing* **21**(2), 141–157.

Blanke, T. and Dunn, S. (2006), The arts and humanities e-science initiative in the uk, *in* Sloot, P. M. A, ed., 'Second IEEE International Conference on e-Science and Grid Computing', IEEE Computer Society, Los Alamitos, Calif. [u.a.], p. 136.

Brockman, W. S. (2001), *Scholarly work in the humanities and the evolving information environment*, Council on Library and Information Resources, Washington DC.

URL: *http://www.clir.org/pubs/reports/pub104/pub104.pdf*

Candela, L., Castelli, D. and Pagano, P. (2013), 'Virtual research environments: An overview and a research agenda', *Data Science Journal* **12**.

Glock, H.-J. (1996), *A Wittgenstein dictionary*, The Blackwell philosopher dictionaries, Blackwell, Oxford and Cambridge, Mass.,.

Gold, M. K., ed. (2012), *Debates in the Digital Humanities*, University of Minnesota Press, Minneapolis.

Gradmann, S. (2010), 'Knowledge = information in context: on the importance of semantic contextualisation in europeana'.

URL: *http://de.scribd.com/doc/32110457/Europeana-White-Paper-1*

Gradmann, S. and Meister, J. C. (2008), 'Digital document and interpretation: re-thinking "text" and scholarship in electronic settings', *Poiesis & Praxis* **5**(2), 139–153.

Grassi, M., Morbidoni, C., Nucci, M., Fonda, S. and Ledda, G. (2012), Pundit: Semantically structured annotations for web contents and digital libraries, *in* A. Mitschick, F. Loizides, L. Predoiu, A. Nürnberger and S. Ross, eds, 'Semantic Digital Archives 2012: Proceedings of the Second International Workshop on Semantic Digital Archives', pp. 49–60.

URL: *http://ceur-ws.org/Vol-912/paper4.pdf*

Hughes, L. (2013), 'Nedimah: Network of digital methods in the arts and humanities. presentation held at at digital humanities luxembourg, december 2013.'.

URL: *http://de.slideshare.net/lorna_hughes/lorna-hughes-12-052013-nedimah-and-ontology-for-dh*

McCarty, W. (2004), Modeling: A study in words and meanings., *in* S. Schreibman, R. Siemens and J. Unsworth, eds, 'A companion to digital humanities', Vol. 26 of *Blackwell Companions to Literature and Culture*, Blackwell, Malden, Mass.

McCarty, W. (2005), *Humanities computing*, Palgrave Macmillan, Houndmills [u.a.].

McCarty, W. and Short, H. (2002), 'Mapping the field: Report of allc meeting held in pisa, april 2002'.

URL: *http://www.allc.org/node/188*

Moulin, C., Nyhan, J., Ciula, A., Kelleher, M., Mittler, E., Tadić, M., Ågren, M., Bozzi, A. and Kuutma, K., eds (2011*a*), *Research Infrastructures in the Digital Humanities*, Vol. 42 of *Science Policy Briefing*.

URL: *http://www.esf.org/fileadmin/Public_documents/Publications/spb42_RI_DigitalHumanities.pdf*

Moulin, C., Nyhan, J., Ciula, A., Kelleher, M., Mittler, E., Tadić, M., Ågren, M., Bozzi, A. and Kuutma, K., eds (2011*b*), *Research Infrastructures in the Digital Humanities: Executive Summary*, Vol. 42 of *Science Policy Briefing*.
URL: *www.esf.org/fileadmin/Public_documents/Publications/spb42_ExecSum.pdf*

Oldman, D., Doerr, M. and Gradmann, S. (tbp), Zen and the art of linked data. new strategies for a semantic web of humanist knowledge. to be published, *in* S. Schreibman, R. Siemens and J. Unsworth, eds, 'A new Companion to Digital Humanities'.

Palmer, C. L., Teffeau, L. C. and Pirmann, C. M. (2009), *Scholarly information practices in the online environment: Themes from the literature and implications for library service development*, OCLC Programs and Research, Dublin, Ohio.
URL: *http://www.oclc.org/content/dam/research/publications/library/2009/2009-02.pdf*

Rockwell, G. (2010), 'As transparent as infrastructure: On the research of cyberinfrastructure in the humanities'.
URL: *http://cnx.org/content/m34315/1.2/*

Terras, M., Nyhan, J. and Vanhoutte, E., eds (2013), *Defining Digital Humanities. A Reader.*, Ashgate, Farnham.

Thaller, M. (2013), 'Praising imperfection: Why editions do not have to befinished'.
URL: *http://www.culingtec.uni-leipzig.de/ESU_C_T/node/292*

Unsworth, J. (2000), 'scholarly primitives: What methods do humanities researchers have in common, and how might our tools reflect this? part of a

symposium on humanities computing: formal methods, experimental practice sponsored by king's college, london, may 13, 2000.'.

URL: *http://people.brandeis.edu/~unsworth/Kings.5-00/primitives.html*

Unsworth, J. and et al. (2006), 'Our cultural commonwealth. report of the american council of learned societies commission on cyberinfrastructure for the humanities and social sciences'.

URL: *http://www.acls.org/programs/cyberinfrastructure/*

Unsworth, J. and Tupman, C. (2012), Interview with john unsworth, april 2011, carried out and transcribed by charlotte tupman, *in* M. Deegan and W. McCarty, eds, 'Collaborative research in the digital humanities', Ashgate, Farnham, pp. 231–240.

Appendix: Scholarly Activities

Note: Where concepts of Scholarly Activities have been essentially reused or remain close the original conceptualisation, references to the original and closest descriptions are provided. The formal scope notes provided here are mostly more exclusive than the original descriptions and reflect our particular interpretation and conceptualisation of the original concepts. Where no reference is given the Scholarly Activity has no appropriate equivalent.

Activity Label	Scope Note
Searching	Direct searching with a well-defined goal (e.g. 'known-item' searches) for specific information or objects of interest which also 'involves deciding where and how to look for information' (cf. Palmer et al., 2009, 9-11).
Discovering	Discovering information or objects of interest through various aids including conversational means. Foraging can be seen as an alternative term which stresses the aspect of discovery in contrast to (direct) searching (cf. Bamboo, 2010, 3-4).
Browsing	Exploratory and open-ended browsing in a body of assembled information such as web pages, library catalogues, archival aids, bookshelves, tables of contents in journals, etc. (cf. Palmer et al., 2009, 13-14)
Probing	Exploratory and investigative strategy employed to find information in unfamiliar domains or topics. May utilise various exploratory means such as database search, archival aids, conversation with domain experts or translating unfamiliar terminology (cf. Palmer et al., 2009, 14-15).
Chaining	Following chains of citations or references either performed as backward chaining (footnote chasing, following references) or forward chaining (citation searching) (cf. Palmer et al., 2009, 11-13).

Monitoring	Keeping constantly and periodically track of developments and news in a field or related to a topic. Essentially an exploratory Activity which might entail other Activities such as chaining, searching, browsing, scanning and reading (cf. Palmer et al., 2009, 29-30).
Reading	Close reading, but might include other kinds and stages of reading such as scanning or systematic skimming, prior to close reading or rereading (cf. Palmer et al., 2009, 19-21).
Contextualising	Adding to the corpus referential structures or referential data by creating relationships between one or more of it elements. Can be seen as more special type of referring (cf. Bamboo, 2010, 5-6).
Translating	Converting and interpreting of new terminology, concepts, theories, methods, etc. for oneself but also for different audiences (cf. Palmer et al., 2009, 31).
Assessing	Determining the quality of an object of interest or information in terms of its relevance, utility, provenance etc. (cf. Palmer et al., 2009, 20-21).
Comparing	Measuring the differences between elements in terms of their structural and conceptual features (cf. Unsworth, 2000).
Filtering	Generating a (temporary) view on the corpus on the basis of one or more criteria. Can also be part of the exploration process (cf. Bamboo, 2010, 4-5).
Sampling	Sampling is a specific subtype of selection in so far as it constitutes a new corpus (the sample) as a subset of the original corpus. Both Selecting and Sampling re-arrange a corpus into a new state or constitute a new one. However, sampling is always performed on an existing corpus (cf. Unsworth, 2000).
Organising	Applying or devising (personal) organisational systems and tools for storing and managing the corpus, its contents or other collections (cf. Palmer et al., 2009, 18-19).

Collecting	Building (personal) collections for current or long-term research including any kind of objects of interest and information (cf. Palmer et al., 2009, 16-19).
Referring	Referencing or linking between two elements, e.g. via a hypertext link or by making a citation (cf. Unsworth, 2000).
Annotating	Adding any kind of notes or markings to any part or element of the corpus (cf. Unsworth, 2000; Bamboo, 2010, 7-8).
Selecting	Adding objects of interest or information to the corpus or removing elements from the corpus based on certain criteria. Selecting modifies an existing corpus by removing and adding elements or constitutes a new corpus by adding the first element to it.
Writing	Proper writing, e.g. of a draft for a journal article or a thesis chapter (cf. Palmer et al., 2009, 21).
Assembling	Putting any kind of elements from the corpus together to form a work which can be shared, published or disseminated. An iterative and continuous process which is based on or may involve other Activities such as writing, reading, sampling etc. (cf. Palmer et al., 2009, 22)
Notetaking	Jotting down thoughts, remarks or notes at any stage of the working and research process and independently from particular objects of interest (cf. Palmer et al., 2009, 10-11).
Illustrating	Visualising an idea, an argument, a relationship or context expressed by text, speech or other visual aids (cf. Unsworth, 2000; Bamboo, 2010, 8).
Sharing	Making (intermediate) research results available to a (selected) audience such as a working group (cf. Bamboo, 2010, 9-11).
Publishing	Making (intermediate) research results available to a wider audience such as the general public (cf. Bamboo, 2010, 9-11).

Disseminating	Making (intermediate) research results available on a more collaborative, continuous and social basis such as attending and speaking at meetings, conferences, scholarly associations and societies (cf. Palmer et al., 2009, 23-25).